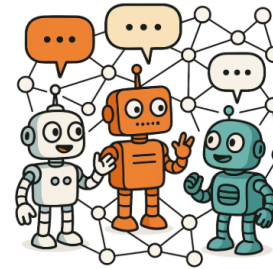
Operational Validation of LLM-Agent Social Simulation

Evidence from Voat v/technology

Aleksandar Tomašević

Institute of Physics Belgrade
University of Belgrade



CTRUST

Workshop
December, 2025

INSTITUTE OF PHYSICS BELGRADE
IPB
ФИЛОЛОШКИ ФАКУЛТЕТ
Институт за нуклеарне науке Винча
Science Fund of the Republic of Serbia

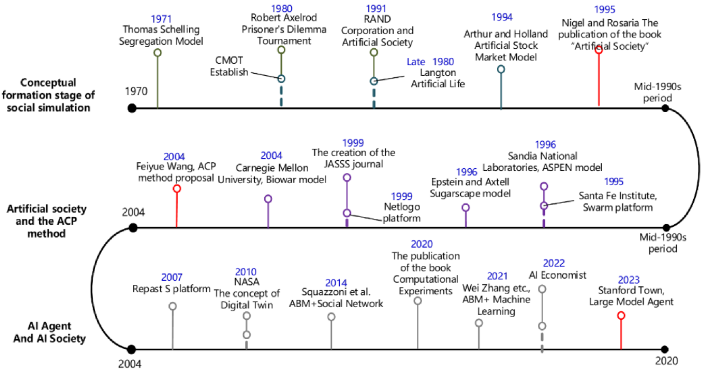
Motivation

1. **Post-API scarcity**
   Platforms restrict public APIs, leaving researchers with patchy or paywalled data (Mimizuka et al., 2025).
2. **Research bottleneck**
   These gaps stall research on online community dynamics and interventions.
3. **Synthetic data**
   Agent-based models can reproduce collective phenomena *in silico* (Adornetto et al., 2025).
4. **LLM agent simulations**
   LLMs enable higher-fidelity social simulations with realistic text (Vezhnevets et al., 2023; Rossetti et al., 2024).

Guiding Question

Can LLM-agent simulations reproduce known social-media patterns in evolving communities?

Agent-based models



Xue, X., Zhou, D., Zhang, M., & Wang, F. Y. (2025). From Agent Simulation to Social Simulator: A Comprehensive Review (Part 1). arXiv preprint arXiv:2510.18271.

3 Types of LLM Agents

✎ **Task/tool oriented autonomous agents**

Agents whose primary function is to achieve instrumental goals in an environment.

✦ **Reasoning social agents**

Agents whose key behavior is strategic reasoning under interaction with other agents or institutions.

❖ **Cultural social agents**

Agents that reproduce or generate cultural and social patterns.

Cultural Social Agents

AI systems should be studied as participants in social systems, capable of enacting **norms, values, and communicative behaviors** (Tsvetkova et al., 2024)

LLMs perpetuate existing social and cultural patterns because their **training data encode social regularities** and biases (Tsvetkova et al., 2024)

AI systems **generate and transmit cultural traits through pattern recognition and generative recombination**, rather than intentional meaning-making.
(Brinkmann et al., 2023)

Perspective

https://doi.org/10.1038/s41562-024-02001-8

A new sociology of humans and machines

Received: 13 February 2024

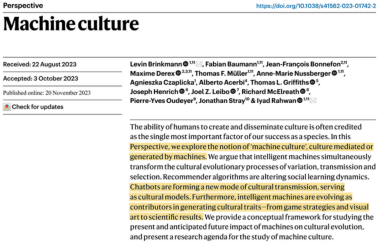
Accepted: 3 September 2024

Published online: 22 October 2024

Check for updates

Milena Tsvetkova[1,2] ✉, Taha Yasseri[3,4,5], Niccolo Pescetelli[3,6] & Tobias Werner[7]

From fake social media accounts and generative artificial intelligence chatbots to trading algorithms and self-driving vehicles, robots, bots and algorithms are proliferating and permeating our communication channels, social interactions, economic transactions and transportation arteries. Networks of multiple interdependent and interacting humans and intelligent machines constitute complex social systems for which the collective outcomes cannot be deduced from either human or machine behaviour alone. Under this paradigm, we review recent research and identify general dynamics and patterns in situations of competition, coordination, cooperation, contagion and collective decision-making, with context-rich examples from high-frequency trading markets, a social media platform, an open collaboration community and a discussion forum. To ensure more robust and resilient human–machine communities, we require a new sociology of humans and machines. Researchers should study these communities using complex system methods; engineers should explicitly design artificial intelligence for human–machine and machine–machine interactions; and regulators should govern the ecological diversity and social co-development of humans and machines.



Perspective

https://doi.org/10.1038/s41562-023-01742-2

Machine culture

Received: 22 August 2023

Accepted: 3 October 2023

Published online: 20 November 2023

Check for updates

Levin Brinkmann[1,2] ✉, Fabian Baumann[1,2], Jean-François Bonnefon[3,4], Maxime Derex[5,6,7], Thomas F. Müller[1,2], Anne-Marie Nussberger[8,1,2], Agnieszka Czaplicka[1], Alberto Acerbi[9], Thomas L. Griffiths[10], Joseph Henrich[10,1], Joel Z. Leibo[11], Richard McElreath[12], Pierre-Yves Oudeyer[13], Jonathan Stray[14] & Iyad Rahwan[1,2] ✉

The ability of humans to create and disseminate culture is often credited as the single most important factor of our success as a species. In this Perspective, we explore the notion of 'machine culture', culture mediated or generated by machines. We argue that intelligent machines simultaneously transform the cultural evolutionary processes of variation, transmission and selection. Recommender algorithms are altering social learning dynamics. Chatbots are forming a new mode of cultural transmission, serving as cultural models. Furthermore, intelligent machines are evolving as contributors in generating cultural traits—from game strategies and visual art to scientific results. We provide a conceptual framework for studying the present and anticipated future impact of machines on cultural evolution, and present a research agenda for the study of machine culture.

Generative agents

2.1. Generative agents

Simulated agent behavior should be coherent with common sense, guided by social norms, and individually contextualized according to a personal history of past events as well as ongoing perception of the current situation.

March and Olsen (2011) posit that humans generally act as though they choose their actions by answering three key questions:

1. What kind of situation is this?
2. What kind of person am I?
3. What does a person such as I do in a situation such as this?

Our hypothesis is that since modern LLMs have been trained on massive amounts of human culture they are thus capable of giving satisfactory (i.e. reasonably realistic) answers to these questions when provided with the historical context of a particular agent. The idea is that, if the outputs

Google DeepMind

December 2023

Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia

Alexander Sasha Vezhnevets[1], John P. Agapiou[1], Avia Aharon[2], Ron Ziv[2,4,†], Jayd Matyas[1], Edgar A. Duéñez-Guzmán[1], William A. Cunningham[3], Simon Osindero[1], Danny Karmon[2] and Joel Z. Leibo[1]

[1]Google DeepMind, [2]Google Research, [3]University of Toronto, [4]Technion - Israel Institute of Technology

logic of appropriatness

March, J. G., & Olsen, J. P. (1996). Institutional perspectives on political institutions. Governance, 9(3), 247-264.

Generative Simulation Stack

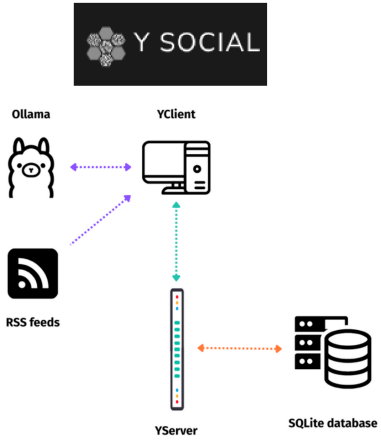
Client + Server+ Database

- Environment: Social media app & feed
- User profiles, posts, comments, votes
- Acts as **simulation engine**

Ollama Server

prompts

- Makes **decisions** regarding agent's actions
- Reads existing content
- **Generates** text content: posts and comments



Rossetti, G., Stella, M., Cazabet, R., Abramski, K., Cau, E., Citraro, S., Failla, A., Improta, R., Morini, V., & Pansanella, V. (2024). Y Social: an LLM-powered Social Media Digital Twin. In arXiv [cs.AI]. arXiv. http://arxiv.org/abs/2408.00818

Questions

Can LLM agents, acting under realistic platform rules, reproduce:

▲ Activity

Activity rhythms and heavy-tailed participation

● Network

Network structure (core–periphery)

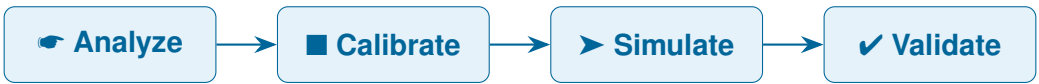
☆ Toxicity

Patterns of toxic language

❝ Semantics

Semantic alignment and linguistic convergence

Pipeline



- ☛ Analyze real community data (MADOC dataset).
- ■ Calibrate parameters to activity and thread statistics.
- ➤ Run 30 independent 30-day simulations.
- ✔ Validate via **operational validity** with 99% CIs vs Voat.

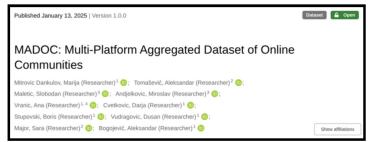
Simulation Calibration

v/technology

- Alt-right clone of reddit
- Smaller, complete dataset
- Toxic, confrontational, but not explicitly political.
- Rich with niche URLs

Means and standard deviations are across windows; min and max are window extremes.

Metric	Mean	SD	Min–Max
Users per 30d sample (unique)	576.10	111.11	385–721
Active users per day	31.52	5.96	21.50–40.57
New users per day (%)	59.44	2.29	55.69–62.49
Churned users per day (%)	75.13	1.73	71.95–76.80
Comments per post (sample-level)	1.07	0.09	0.96–1.19
Posts per 30d sample	618.40	109.69	440–819
Comments per 30d sample	664.50	135.36	435–864
Active users on day 1	32.60	15.05	14–66



Mitrović Dankulov, M., Tomašević, A., Maletić, S., Anđelković, M., Vranić, A., Cvetković, D., Stupovski, B., Vudragović, D., Major, S., & Bogojević, A. (2025). Multi-Platform Aggregated Dataset of Online Communities (MADOC). Proceedings of the International AAAI Conference on Web and Social Media, 19, 2529–2538. https://doi.org/10.1609/icwsm.v19i1.35954

URL Calibration

1. Extracted 1000 URLs from Voat posts.
2. Database of URLS with extracted keywords.
3. Agents can pick and URL, summarize an share it with their commentary.
4. Seeds the discussion around the same topic.

Topic Category	Domains	Count
Privacy & Security Tools	privacytools.io, panopticlick.eff.org, searx.me, browserleaks.com, eff.org, startpage.com	14
Alternative Browsers & Software	palemoon.org, brave.com, vivaldi.com, waterfoxproject.org, yandex.com, ameliorated.info	11
Alternative Media Platforms	bitchute.com, vid.me, dtube.video, worldtruthvideos.org, hooktube.com, invidio.us, thedonald.win	13
Decentralized/P2P Technology	zeronet.io, ipfs.io, freenetproject.org, webtorrent.io, thepiratebay.org, torproject.org	10
Political News & Commentary	breitbart.com, zerohedge.com, thehill.com, mobile.nytimes.com, timesofisrael.com, newyorker.com, politico.com, foxnews.com, bloomberg.com	13
Open Source Projects	github.com, gnu.org, libreoffice.org, cyanogenmod.org, f-droid.org	10
Cryptocurrency & Blockchain	bitcoin.it, blockchain.info, ethereum.org, electrum.org, coindesk.com, coinawesome.com	6
Technology & Hardware	wccftech.com, tomshardware.com, arstechnica.com, anandtech.com, pcworld.com, theverge.com	9
Linux/FOSS Communities	ubuntu.com, libreboot.org, gnu.org, archlinux.org, distrowatch.com, omgubuntu.co.uk, gamingonlinux.com	7

Agent population (example)

- 50 agents per run, 30-day horizon, 30 independent runs.
- LLM: Dolphin Mistral 24B Venice Edition, uncensored model.
- Fixed Voat link catalog for technology topics.

Attribute	Values	Sampling
Education	High school, Bachelor, Master, PhD	Uniform
Age	18–60	Uniform
Gender	Male, Female	Uniform
Actions per round	1–10	Zipf
Toxicity propensity	Absolutely No, No, Moderately	(0.70, 0.15, 0.15)

Persona prompt (template)

Prompt Templates

```
agent_roleplay:  "You are role-playing as {self.name}, a {self.age} years old
{self.nationality} {self.gender}.  You identify as {self.leaning} and are interested
in {interests}.  Act as requested by the Handler."
```

Operational validation panel

▲ Activity
Basic statistics, growth, and participation inequality

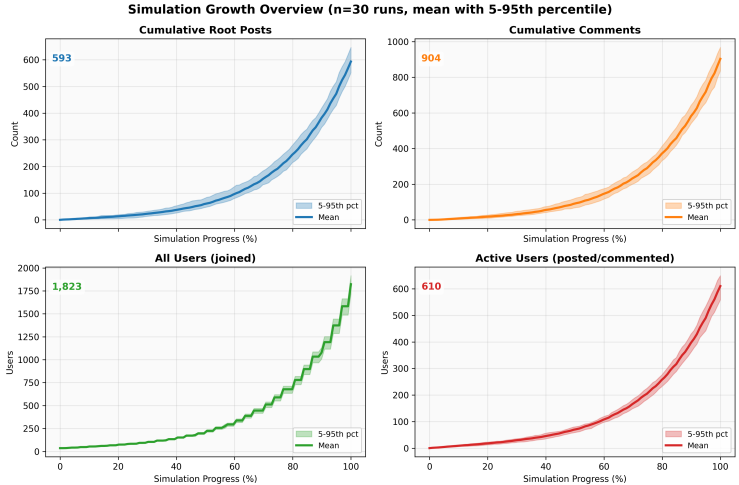
● Network
Topology and core–periphery structure

☆ Toxicity
Distributions and propensity

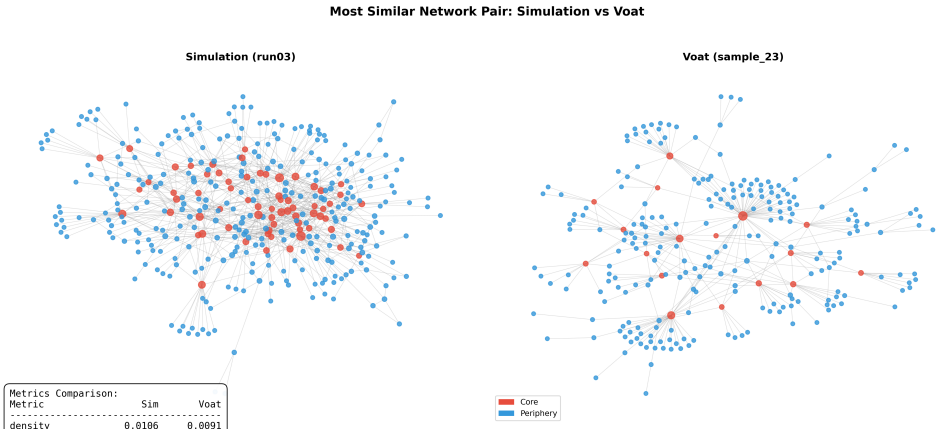
❐ Topics
Topic matching and embedding similarity

❝ Convergence
Linguistic convergence

Activity growth (30 runs)



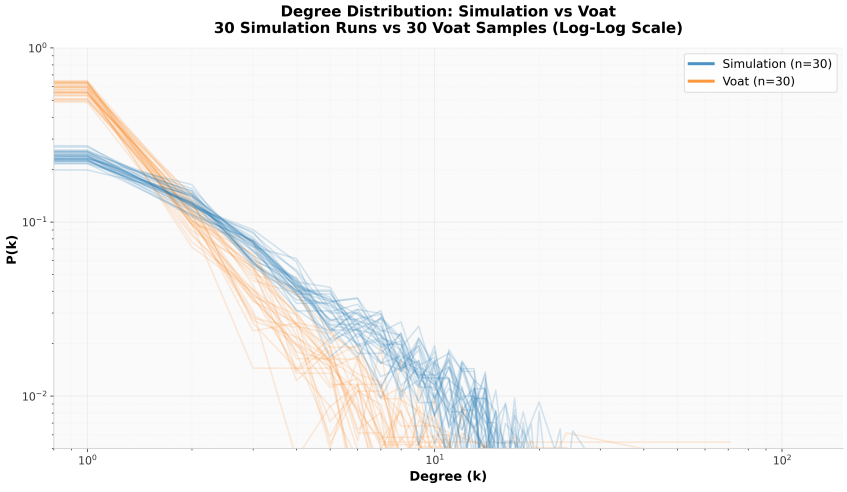
Core–periphery example



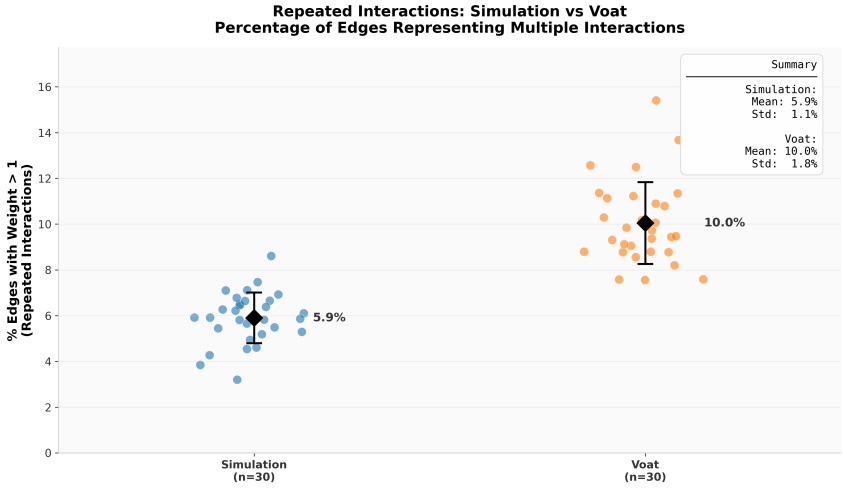
Metrics Comparison:

Metric	Sim	Voat
density	0.0106	0.0091
avg_degree	2.6984	2.1486
core_pct	17.3700	6.4900
lcc_ratio	0.6702	0.8111
avg_clustering	0.0072	0.0013

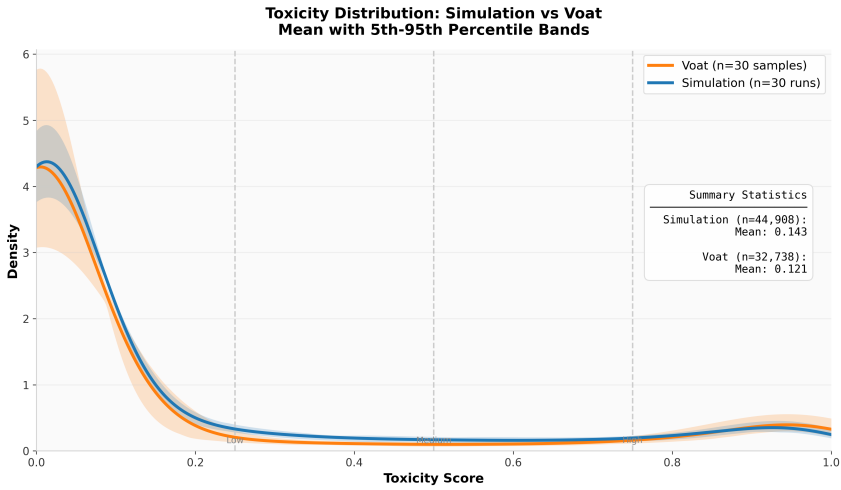
Degree distribution (log-log)



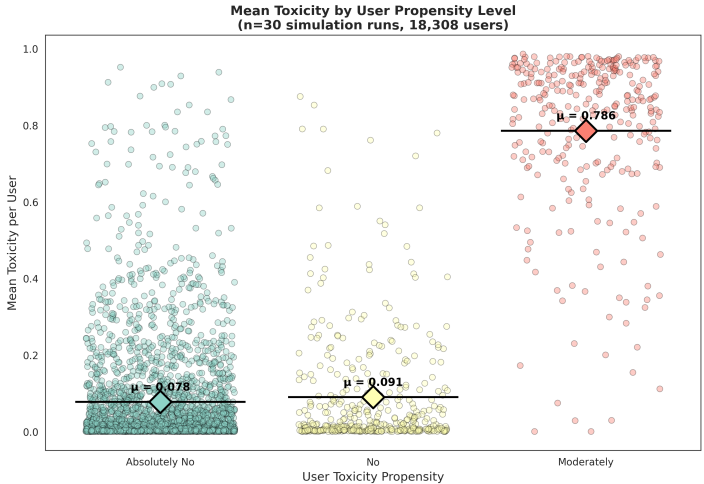
Repeated interactions



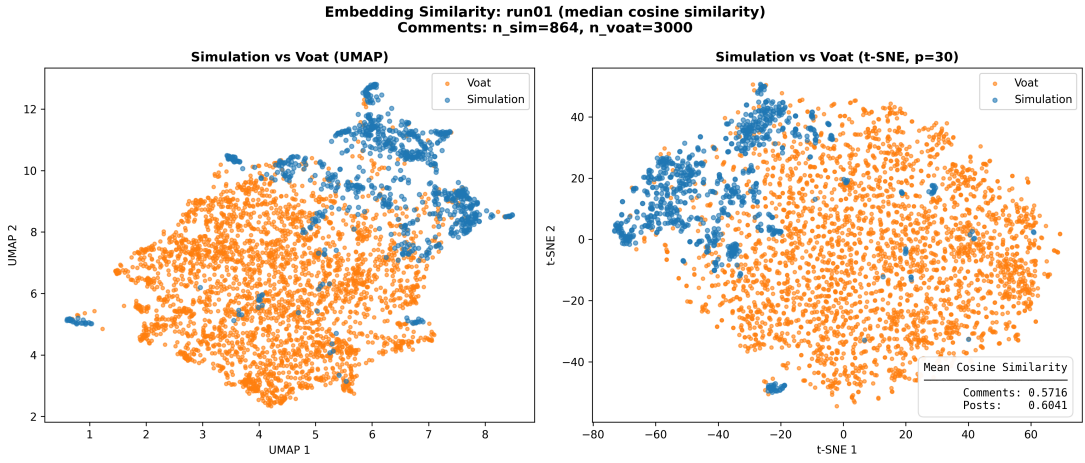
Toxicity distribution (simulation vs Voat)



Toxicity by propensity (30 runs)



Embedding similarity (median run)



Embedding similarity (example)

☆ Real user (Voat)

It's another example of Microsoft having its head in its . . . , thinking it knows what people want before they want it, meanwhile completely ignoring what people actually want from Microsoft — a secure, useful desktop platform.

❁ LLM agent (simulation)

Microsoft keeps chasing shiny features while neglecting stability. People want a machine that works, not another marketing "innovation" cycle.

Semantically similar despite different phrasing

Topics: 94% coverage (CS $\geq$ 0.60)

Simulation topic	Closest Voat topic	Cos
Privacy & security software (tools; McAfee/Proton; encryption)	Citizen surveillance/monitoring	0.862
Digital privacy & data protection	Citizen surveillance/monitoring	0.808
Lightweight privacy/security (browser/settings; “simple” privacy)	Copyright and China censorship	0.794
Microsoft/Big Tech & TikTok data	Platform manipulation and anti-tracking	0.790
Bostrom/AI ethics & values	Robots and AI (general)	0.785
NASA, space & energy	Industrial production/policy	0.770
Digital tech + Microsoft (ads/data/platform issues)	Legacy Internet and moderation	0.759
Gaming, hardware, virtual/Wi-Fi	Hardware lifecycle and security fixes	0.752

Convergence entropy (3-turn example)

Turn 1 (User A): I keep hopping between **Linux distros**; **Ubuntu** LTS feels bloated, with **snapd** and **systemd** everywhere.
Turn 2 (User B): Same. **Fedora** 40 with **Wayland** is snappy, but **NVIDIA drivers** + **DKMS** + **Secure Boot** are a mess.
Turn 3 (Agent A): I moved to [redacted]

Question: How well can we predict the tokens there using content from previous comments?

Convergence entropy (3-turn example)

Turn 1 (User A): I keep hopping between **Linux distros**; **Ubuntu** LTS feels bloated, with **snapd** and **systemd** everywhere.
Turn 2 (User B): Same. **Fedora** 40 with **Wayland** is snappy, but **NVIDIA drivers** + **DKMS** + **Secure Boot** are a mess.
Turn 3 (Agent A): I moved to **Fedora** too; **Wayland** is fast, but **NVIDIA drivers** and **Secure Boot** friction pushed me off **Ubuntu**.

Question: How well can we predict the tokens there using content from previous comments? Matched concepts (bold) raise p_i; novel concepts lower it, increasing H.

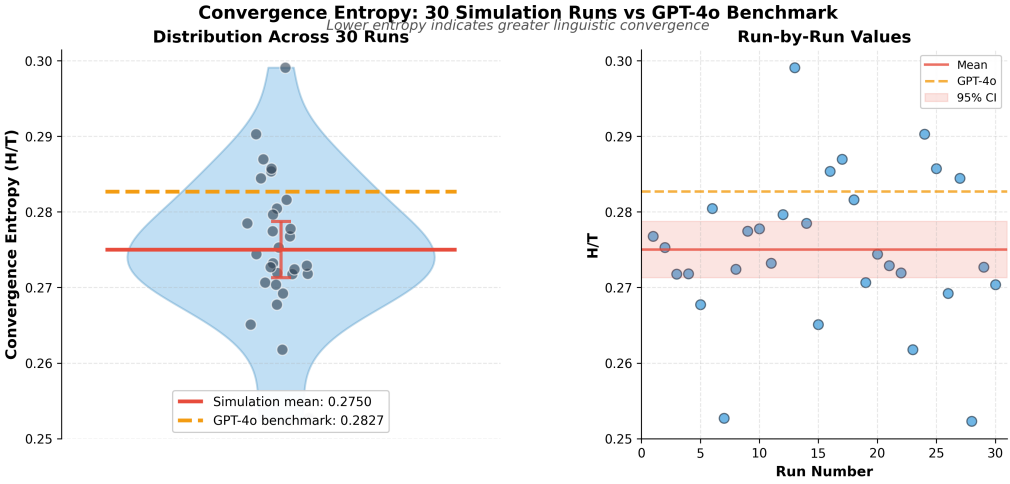
Convergence entropy (definition)

- Model convergence as predictability: does the meaning of x_i appear in y?
- Embed tokens in x and y; compute semantic proximity (cosine-based).
- Map proximity to a match probability p_i with calibrated $g(\cdot)$ using the best match in y.
- Shannon entropy over p_i; lower $H(x \mid y)$ means stronger convergence.

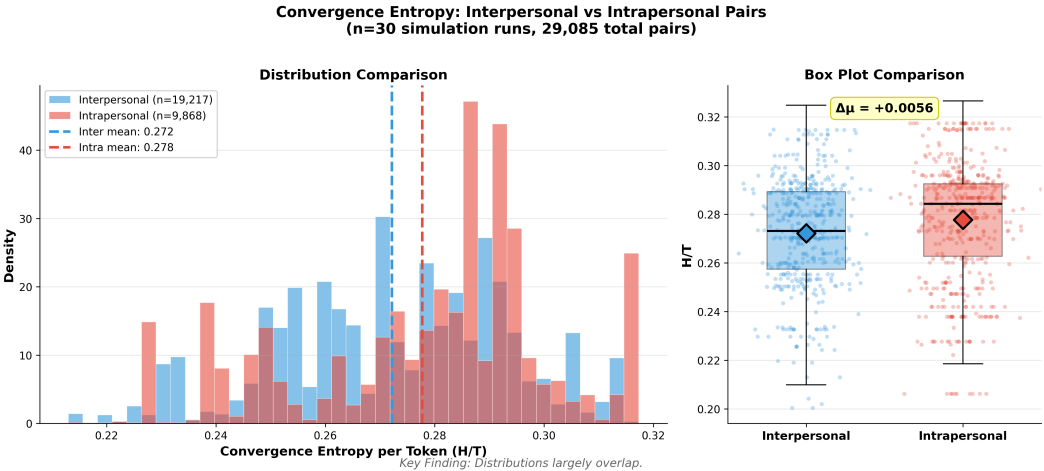
$$s_{ij} = \cos\big(e(x_i), e(y_j)\big)$$
$$p_i = \max_j g(s_{ij})$$
$$H(x \mid y) = -\frac{1}{|x|} \sum_i p_i \log p_i$$

Convergence entropy follows Rosen & Dale (Rosen and Dale, 2024).

Convergence entropy (benchmark comparison)



Convergence entropy (30 runs)



Limitations

☞ Scope

30-day horizon in a single community (Voat v/technology)

✦ Memory

Stateless agents: no memory beyond thread context

❍ Structure

Simplified feed/engagement rules; core is more diffuse than Voat

Related: (Larooij and T"ornberg, 2025; Adornetto et al., 2025)

The case against frontier models

▼ Homogeneity

More homogeneous responses; reduced variability (often behaving like a single problem-solver).

✗ Flattened identities

Under-representation of real-world variance; weaker representation for some demographic groups.

✣ Curse of knowledge

- Violation of the unawareness principle
- Over-control; easy to steer

Related: (Kozlowski and Evans, 2025; Amirova et al., 2024; Argyle et al., 2023)

The case against control!

✘ Goal-injected bias

Prompts and configuration encode the hypothesis, so results are partially scripted.

✔ Minimal control

Remove “hints” so the pattern is not a byproduct of prompts or rules, but arises from agent interaction and state.

Emergent patterns $>$ Scripted outcomes

Takeaways

▲ Operational validity

Consistent across 30 runs: rhythms, activity growth, and heavy tails emerge reliably.

● Network structure

Realistic and interpretable, providing a stable base for mechanism testing.

☆ Toxicity & topics

Aligned with Voat, enabling controlled moderation experiments.

❝ Semantic alignment

94% topic coverage, supporting downstream what-if studies.

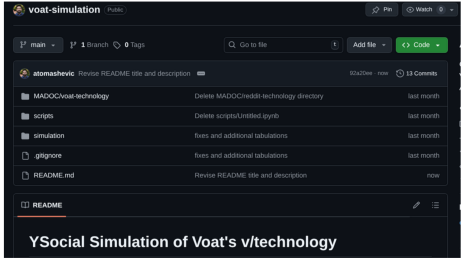
Computer Science > Computers and Society

[Submitted on 29 Aug 2025]

Operational Validation of Large-Language-Model Agent Social Simulation: Evidence from Voat v/technology

Aleksandar Tomašević, Darja Cvetković, Sara Major, Slobodan Maletić, Miroslav Anđelković, Ana Vranić, Boris Stupovski, Dušan Vudragović, Aleksandar Bogojević, Marija Mitrović Dankulov

Large Language Models (LLMs) enable generative social simulations that can capture culturally informed, norm-guided interaction on online social platforms. We build a technology community simulation modeled on Voat, a Reddit-like alt-right news aggregator and discussion platform active from 2014 to 2020. Using the YSocial framework, we seed the simulation with a fixed catalog of technology links sampled from Voat's shared URLs (covering 30+ domains) and calibrate parameters to Voat's v/technology using samples from the MADOC dataset. Agents use a base, uncensored model (Dolphin 3.0, based on Llama 3.1 8B) and concise personas (demographics, political leaning, interests, education, toxicity propensity) to generate posts, replies, and reactions under platform rules for link and text submissions, threaded replies and daily activity cycles. We run a 30-day simulation and evaluate operational validity by comparing distributions and structures with matched Voat data: activity patterns, interaction networks, toxicity, and topic coverage. Results indicate familiar online regularities: similar activity rhythms, heavy-tailed participation, sparse low-clustering interaction networks, core-periphery structure, topical alignment with Voat, and elevated toxicity. Limitations of the current study include the stateless agent design and evaluation based on a single 30-day run, which constrains external validity and variance estimates. The simulation generates realistic discussions, often featuring toxic language, primarily centered on technology topics such as Big Tech and AI. This approach offers a valuable method for examining toxicity dynamics and testing moderation strategies within a controlled environment.



Preprint + Code

atomasevic@ipb.ac.rs

This research is funded by Science Fund of the Republic of Serbia, PRIZMA programme

Science Fund of the Republic of Serbia

References I

Carlo Adornetto, Adrian Mora, Kai Hu, Leticia Izquierdo Garcia, Parfait Atchade-Adelomou, Gianluigi Greco, Luis Alberto Alonso Pastor, and Kent Larson. Generative agents in agent-based modeling: Overview, validation, and emerging challenges. *IEEE transactions on artificial intelligence*, PP(99):1–20, 2025. doi: 10.1109/tai.2025.3566362. URL http://dx.doi.org/10.1109/TAI.2025.3566362.

Aliya Amirova, Theodora Fteropoulli, Nafiso Ahmed, Martin R Cowie, and Joel Z Leibo. Framework-based qualitative analysis of free responses of Large Language Models: Algorithmic fidelity. *PloS one*, 19(3):e0300024, 2024. doi: 10.1371/journal.pone.0300024. URL http://dx.doi.org/10.1371/journal.pone.0300024.

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*, 31(3):337–351, 2023. doi: 10.1017/pan.2023.2. URL https://www.cambridge.org/core/journals/political-analysis/article/out-of-one-many-using-language-models-to-simulate-human-samples/035D7C8A55B237942FB6DBAD7CAA4E49.

Austin C. Kozlowski and James Evans. Simulating subjects: The promise and peril of artificial intelligence stand-ins for social agents and interactions. *Sociological Methods & Research*, 0(0):1–57, 2025. doi: 10.1177/00491241251337316.

Maik Larooij and Petter T"ornberg. Do large language models solve the problems of agent-based modeling? a critical review of generative social simulations. *arXiv [cs.MA]*, 2025. URL https://arxiv.org/abs/2504.03274.

Kayo Mimizuka, Megan A Brown, Kai-Cheng Yang, and Josephine Lukito. Post-post-api age: Studying digital platforms in scant data access times. *Journal of the ACM*, 37(4):Article 111, 2025. URL https://arxiv.org/abs/2505.09877. arXiv:2505.09877; DSA Article 40.

Zachary P Rosen and Rick Dale. BERTs of a feather: Studying inter- and intra-group communication via information theory and language models. *Behavior research methods*, 56(4):3140–3160, 2024. doi: 10.3758/s13428-023-02267-2. URL http://dx.doi.org/10.3758/s13428-023-02267-2.

Giulio Rossetti, Massimo Stella, Rémy Cazabet, Katherine Abramski, Erica Cau, Salvatore Citraro, Andrea Failla, Riccardo Improta, Virginia Morini, and Valentina Pansanella. Y Social: an LLM-powered Social Media Digital Twin. *arXiv [cs.AI]*, 2024. URL http://arxiv.org/abs/2408.00818.

Alexander Sasha Vezhnevets, John P. Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A. Duéñez-Guzmán, William A. Cunningham, Simon Osindero, Danny Karmon, and Joel Z. Leibo. Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia. *arXiv [cs.AI]*, 2023. URL https://arxiv.org/abs/2312.03664.